Supporting Information

Insights Into the Origin of Life: Did it Begin from HCN and H_2O?

Tamal Das,[a,b] Siddharth Ghule[a,b] and Kumar Vanka[a,b,*]

[a]Physical and Material Chemistry Division, CSIR-National Chemical Laboratory (CSIR-NCL), Dr. Homi Bhabha Road, Pashan, Pune 411008, India

[b]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

*Email: k.vanka@ncl.res.in

Table of Contents

Computational Details

Quantum Mechanical Calculations: In order to determine the reaction free energy (ΔG) and energy barriers ($\Delta G^{\#}$), we have done the minimum energy pathway (MEP) search by full quantum mechanical calculations, including zero point energy, internal energy, and entropic contributions, with the temperature taken to be 298.15 K. All the calculations for the structures reported have been done using density functional theory (DFT). Geometry optimizations and transition state search calculations were carried out with the Turbomole 7.0 software package[1] using the TZVP basis set[2] and the B3LYP three parameter hybrid density functional.[3] Dispersion corrections (D3) were included in all the calculations. Solvent corrections were included with the dielectric continuum solvent model COSMO[4] with $\varepsilon = 80.0$. Furthermore, in order to make our data more reliable and also to refine the energies, the single point energy calculation of all the transition states and connecting reactants, intermediates and products were then done at the RI-CC2[5]/TZVP+COSMO(ε=80.0) and RI-MP2[5]/TZVP+COSMO(ε=80.0) level of theory and the corresponding values have been reported in the main manuscript (Figures 3, 4 and 5). The trends for the ΔG and the $\Delta G^{\#}$ values were seen to match for the calculations done at both the levels of theory. In addition to the calculations done with Turbomole 7.0, we have also optimized all the transition states, corresponding reactants and products with the Gaussian09 software package[6] at the B3LYP/6-311++g(d,p)[7] level of theory (with DFT-D2,[8-12] a general, empirical dispersion correction proposed by Stefan Grimme for DFT calculations). Furthermore, we have also done all the calculations with the M06-2X hybrid fuctional[13-15] and the 6-311++g(d,p) basis set. In all the Gaussian09 calculations, we have modeled the solvent with the polarizable continuum model (PCM),[16] with water (ε =80.0) as the solvent. These values are shown in Figures S4, S6, S11 in this SI file. Therefore, we have employed four different levels of theory for the QM calculations: B3LYP-D3/TZVP+COSMO(ε=80.0)//RI-CC2/TZVP+COSMO(ε=80.0) and B3LYP-D3/TZVP+ COSMO(ε=80.0)//RI-MP2/TZVP+COSMO(ε=80.0) in Turbomole 7.0, as well as B3LYP-D2/6-311++g(d,p)+PCM (ε=80.0) and M06-2X/6-311++g(d,p)+PCM (ε=80.0) in Gaussian09 for refining the reaction free energies and barrier heights. For some pathways leading to products that do not form part of the central story, calculations at only the B3LYP-D3/TZVP+COSMO(ε=80.0) level of theory with Turbomole 7.0, for both geometry optimizations and energy calculations, have been done, and the results have been shown in Figures S5, S7, S10, S12 in the SI file. Frequency calculations were performed for all the stationary points to confirm them as a local minima or transition state structures. We have

further done intrinsic reaction coordinate (IRC) calculations to confirm that the obtained transition states connect with the correct reactants and products.

The calculation of the translational entropy in standard software involves assumptions about the volume that may be inaccurate. The translational entropy term can be corrected by a free volume correction introduced by Mammen and co-workers.[17] Based on the Sackur−Tetrode equation, the free volume model describes the translational entropy of molecules in the solution ($\Delta S_{trans}(sol)$); and provides physically intuitive corrections for translational entropy values. In the free volume model, it has been assumed that the volume available to the molecule in solution is lower than the total volume, and this "free volume" is determined by the equation:

$$V_{free} = C_{free}\left(\sqrt[3]{\frac{10^{27}}{[X]N_0}} - \sqrt[3]{V_{molec}}\right)^3$$

Here, V_{molec} is the molecular volume, [X] is the concentration of molecules (mol/L) in solution, and N_0 is the Avogadro number. The translational entropy can be obtained after considering the free volume correction, and inserting the value of V_{free} in the Sackur−Tetrode equation. The total entropy is then calculated by adding the corrected translational entropy and the entropic contributions from the rotational and vibrational components. In our calculations, we have taken C_{free} to be 8.0 for the hard sphere, and [X] = 55.5 mol/lt.

AINR Spherical Boundary Conditions: Spherical boundary conditions were applied to prevent the molecules from flying away, a phenomenon known as the "evaporation" event. The spherical boundary conditions were provided in the form of a sum of two harmonic terms. The molecules were restricted to move inside a spherical volume by a boundary potential, with a time-dependent component :

$$V(r, t)= f(t)U(r,r_1,k_1) +(1− f(t)) U(r,r_2,k_2)$$

$$U(r,r_0,k) = mk/2\ (r – r_0)^2\theta(r – r_0);\ f(t) = \theta(\lfloor t/T \rfloor − t/T + \tau/T)$$

where k_1 = 1.0 kcal mol^{-1}Å^{-2}, r_1 = 10.0 Å, k_2 = 0.5 kcal mol^{-1}Å^{-2}, r_2 = 3.5 Å, τ = 1.7 ps, T = 2.0 ps, ⌊⌋ is the floor function and θ is the heaviside step function. The function f(t) is a rectangular wave that oscillates between one (duration τ) and zero (duration T–τ), and U(r, r_0, k) is a radial potential that is zero inside the prescribed radius r_0 and harmonic outside. The force constant is multiplied by the atomic mass (in a.m.u) such that all the atoms at the same radial coordinate have attained equal acceleration. The rectangular waveform switches the restraint potential

between U(r, r_1, k_1) and U(r, r_2, k_2), which forces the atoms with a radial position 10.0 Å to 3.5 Å towards the centre of the sphere and allows them to collide. When the sphere is expanded again, the molecules present in the smaller volume diffuse rapidly (because of the high simulation temperature) to occupy the larger volume. Due to the repeating compression and expansion of spherical volume, the molecules collide and relax. Therefore, throughout the simulation, new molecules are formed and then break again to form other new molecules. We have run simulations with 93 atoms (16 H_2O + 15 HCN) for a total time of 750.0 ps with a timestep of 0.5 fs.

Optimizations of the *ab initio* Nanoreactor (AINR) Simulations: We have done several AINR simulations for optimizing the initial conditions of the simulations, based on the different parameters that can affect the results. The parameters are (i) the ratio of the reactant species, (ii) the number of molecules taken in the simulation box, (iii) spherical boundary conditions, (iv) temperature, and (v) the total time of the AIMD simulations.

We now briefly discuss the results obtained from our several simulations based on the tested nanoreactor parameters:

(i) We have done the AINR simulation with a 1:2 mixture of HCN (9 molecules) and H_2O (18 molecules), while setting the other parameters to be k_1 = 1.0 kcal $mol^{-1}Å^{-2}$ (the force constant at the outer boundary), r_1 = 8.5 Å, k_2 = 0.5 kcal $mol^{-1}Å^{-2}$ (the force constant at the inner boundary), r_2 = 3.0 Å, τ = 1.7 ps, Total time between collisions = 2.0 ps. In this simulation, we have found the initial hydrolyzed products of HCN such as formamide, formic acid and formaldehyde, but not the other intermediates species or the RNA and protein precursors. The mechanistic pathways for the formation of the hydrolyzed products are similar to those that we have found in our main AINR simulation. The total time evolution for this simulation is ~700ps.

(ii) In another AINR simulation, we have taken a 2:1 mixture of HCN (18 molecules) and H_2O (9 molecules), keeping all other parameters the same as (i) and running for ~700ps. In this case, instead of hydrolyzed products, we have found more oligomeric products of HCN. Very few formamide molecules were formed, and due to the lack of water molecules, the system did not further lead to the formation of the desired RNA and protein precursor molecules.

(iii) In our simulations, the source of carbon and nitrogen is HCN, and the oxygen source is water. This has influenced our decision to take almost a 1:1 ratio of HCN and H_2O, so that it can maximise the interaction between the two different moieties. In order to check the validity of this approach, we have done two different simulations, where we have taken (11 HCN +13

H_2O) and (15 HCN +16 H_2O) mixtures and run the AINR simulations for ~750ps. In both these cases, we have come up with the desired RNA and protein precursors. In our current manuscript, we have reported the results obtained from the (15 HCN +16 H_2O) mixture AINR simulation. Therefore, the results indicate that if one takes a similar number of HCN and H_2O molecules, i.e. in about a 1:1 ratio, it will maximise the probability of getting the desired final products.

(iv) Another important parameter in the AINR simulations is the spherical boundary condition, which we have discussed in the "**AINR Spherical Boundary Conditions**" subsection in the Computational Details section here in the SI. Optimizing the boundary in the AINR is a trial and error process. We had to fix the boundary in such a way so that the collision between the molecules in the inner sphere would be effective and the molecules would get enough space to relax in the outer sphere. Also, the point to be noted is that the collisions should not be at such a high velocity that the molecules breaks into it elemental form.

(v) Temperature is another important parameter in the AINR simulations. The temperature that we have used in our simulations (2000 K) is not the actual reaction temperature at which the reactions would occur. The reason that we have provided such a high temperature is to avoid noncovalent interactions such as hydrogen bonding in our AINR simulation and also to provide enough kinetic energy to the molecules so that they could collide with each other and cross the activation barriers, leading to the products. We reiterate that the purpose of the AINR is to act as a tool for discovery of chemical reactions, the feasibility of which could then be determined with careful, high level QM (DFT, MP2 and CC2) studies of the thermodynamics and kinetics of the discovered reactions. Therefore the parameters (ratio of the reactants, spherical boundary conditions and temperature) that would lead to the best possibility of discovering new processes have to be employed, regardless of whether they necessarily represent the actual experimental conditions or not.

(vi) The goal of the AINR simulations is to find new reactions and mechanistic pathways, and not to equilibrate the systems. We have run most of the simulations at a ~1ns timescale, which sufficed to yield different interesting intermediates and products, as well as the corresponding mechanistic pathways.

Analysis of the Output from the *ab initio* Nanoreactor: After the AIMD run, we have analysed the simulation trajectories. What this involves is the identification of new molecules and the pathways of their formation. This was done by using data analysis and visualization with the Python libraries NetworkX,[18] Numpy[19] and Graphviz.[20] The two-state hidden Markov

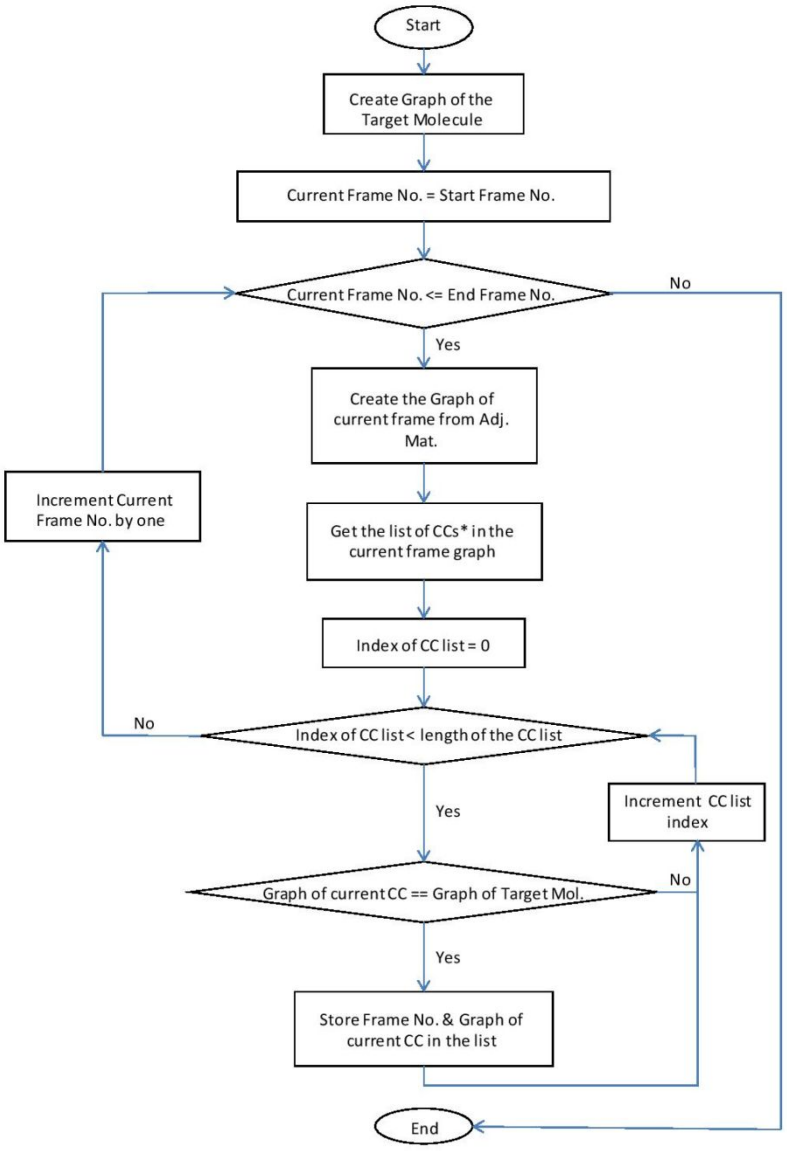
model[21] (HMM) was employed for this purpose. The description of the HMM model is provided below. We would like to point out, however, that we did not observe any significant improvement in the trajectory analysis tree by applying the HMM model. Therefore, we have relied more on our Python code as well as on the manual visualization of the simulation trajectories in the Molden[22] and VMD[23] softwares, for determining the best possible routes to the formation of the intermediate molecules observed in the nanoreactor.

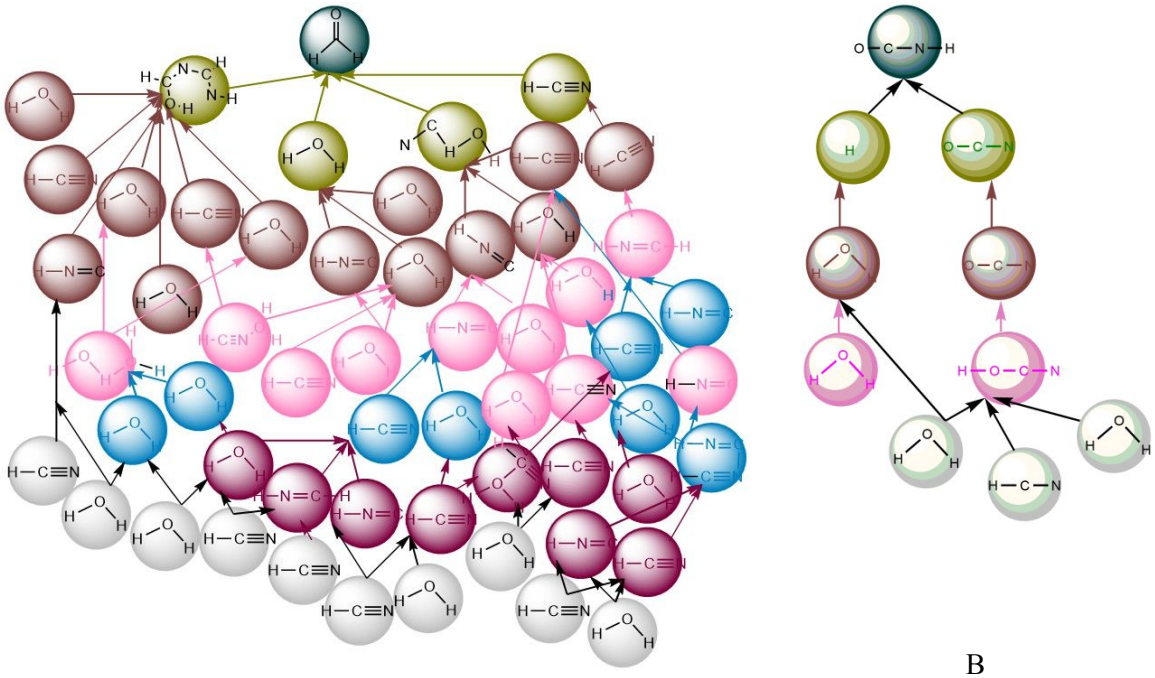
The Hidden Markov Model (HMM):

The hidden Markov model (HMM) is a tool for representing the probability distribution over a sequence of observations. The HMM has two important properties. First, the observations are generated by a process whose states are hidden from the observer. Second, it assumes that the state of this hidden process satisfies the Markov property, i.e., when predicting the future, the past does not matter, only the present. The HMM allows us to predict the most probable sequence of hidden states for the given sequence of observations. It is specified by transition and emission probabilities.

For the current data, produced from the *ab initio* nanoreactor simulations, the hidden and observed states are the same, i.e. 0 and 1 (1 = bond and 0 = no bond between a pair of atoms). These states are elements of a connectivity matrix. We obtain the sequence of observed states from the simulations and predict the hidden states from the HMM. We have employed the *Viterbi* algorithm[24] to find the most probable sequence of the hidden states. The connectivity matrix constructed from these hidden states is supposed to give us improved connectivity between the atoms. The observed sequence from the whole cycle (collision + non-collision steps) was provided as input to the model. As the collision steps are most complicated to analyze, we have employed the HMM to construct the connectivity matrix of only the collision steps. The transition and emission probabilities were obtained from the original *ab initio* nanoreactor paper by Martinez and co-workers.[25]

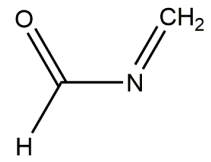
Flowchart of the Implementation of the Hidden Markov Model (HMM)


Start
Create Graph of the Target Molecule
Current Frame No. = Start Frame No.
Current Frame No. <= End Frame No.
No
Yes
Create the Graph of current frame from Adj. Mat.
Increment Current Frame No. by one
Get the list of CCs* in the current frame graph
Index of CC list = 0
No
Index of CC list < length of the CC list
Increment CC list index
Yes
No
Graph of current CC == Graph of Target Mol.
Yes
Store Frame No. & Graph of current CC in the list
End
*CC=Connected Components

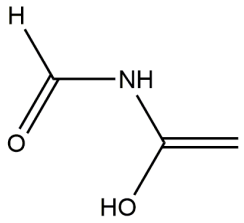


A B

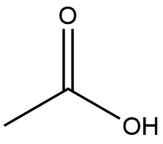
Figure S1. The representation of the complex connectivity graph for the formation of formaldehyde (A), and the simplest connectivity graph for the formation of isocyanic acid (B), starting from HCN and H_2O. How the connectivity graph changes after each collision is represented by different colours of the sphere: deep blue represents the target molecule; green, brown, pink are the intermediate species and at the bottom, white represents the starting molecules. For further understanding, the reader is also encouraged to look at the methodology outlined in the paper by Martinez and co-workers.[25]



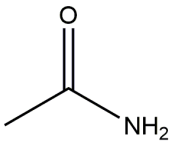
N-methyleneformamide



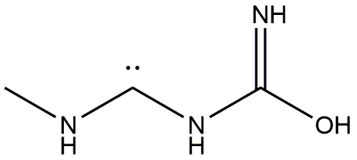
N-(1-hydroxyvinyl)formamide



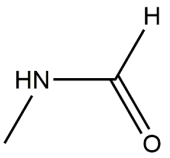
acetic acid



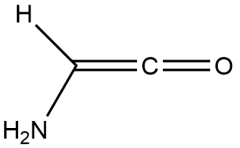
acetamide



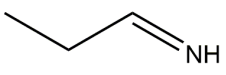
(*E*)-*N*-(((hydroxy(imino)methyl)amino)methanidylene)methanaminium



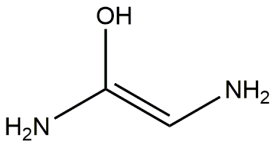
N-methylformamide



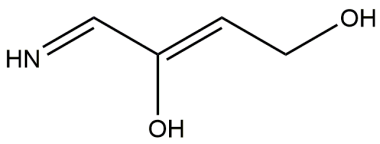
2-aminoethen-1-one



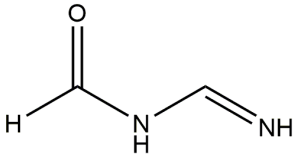
propan-1-imine



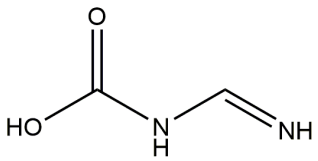
(*Z*)-1,2-diaminoethen-1-ol



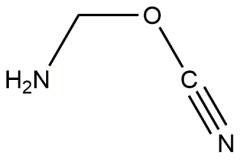
(*Z*)-4-iminobut-2-ene-1,3-diol



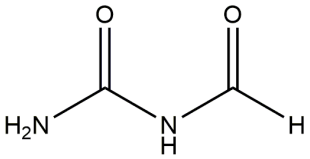
N-(iminomethyl)formamide



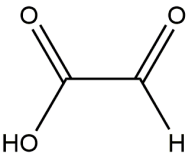
(iminomethyl)carbamic acid



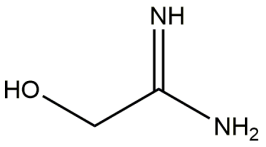
cyanatomethanamine



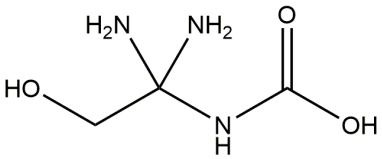
N-carbamoylformamide



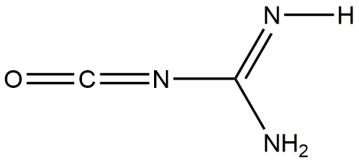
2-oxoacetic acid



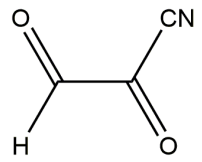
2-hydroxyacetimidamide



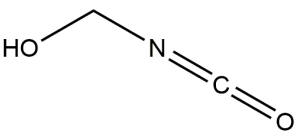
(1,1-diamino-2-hydroxyethyl)carbamic acid



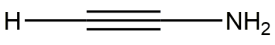
carbamimidoyl isocyanate



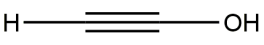
2-oxoacetyl cyanide



isocyanatomethanol



ethynamine



ethynol

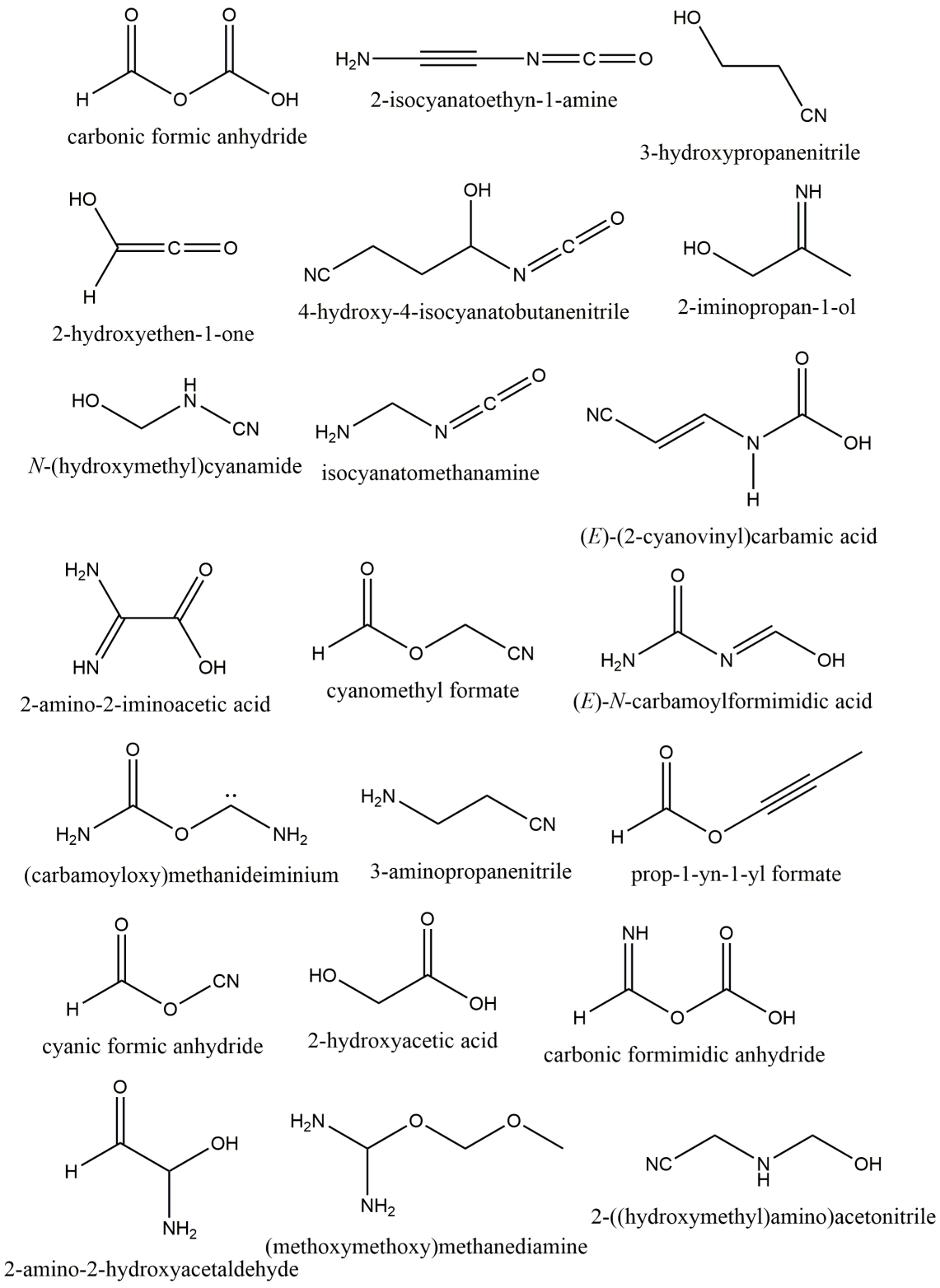


Figure S2. A selection of the products that were discovered from the *ab initio* nanoreactor simulations, including ribonucleotide and amino acid precursors and other intermediate compounds discussed in the main manuscript. Here, we have shown that apart from the important precursors for RNA and protein, a lot of diverse acyclic organic compounds were also formed during the simulations.

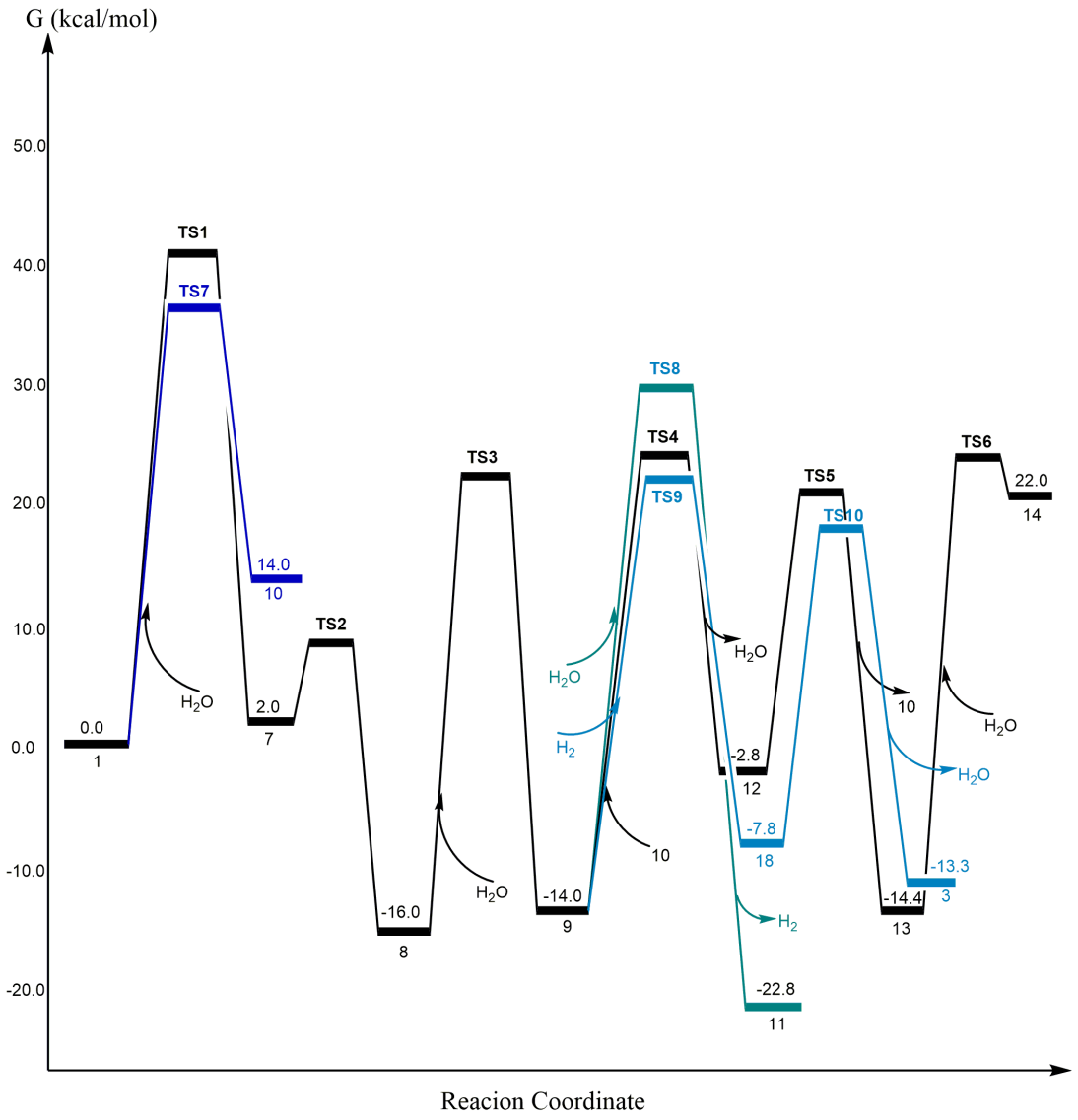


Figure S3. The free energy profile for the formation of intermediate molecules of the RNA and protein precursors starting from HCN and water molecules. The relative energies of the reactants and products of each elementary step have been represented with respect to HCN and the barrier has been calculated from the reactant species for each elementary step reaction. The values (in kcal/mol) have been obtained at the B3LYP-D2/6-311++g(d,p)+PCM(Ɛ=80.0) level of theory by Gaussian09 software package.[6]

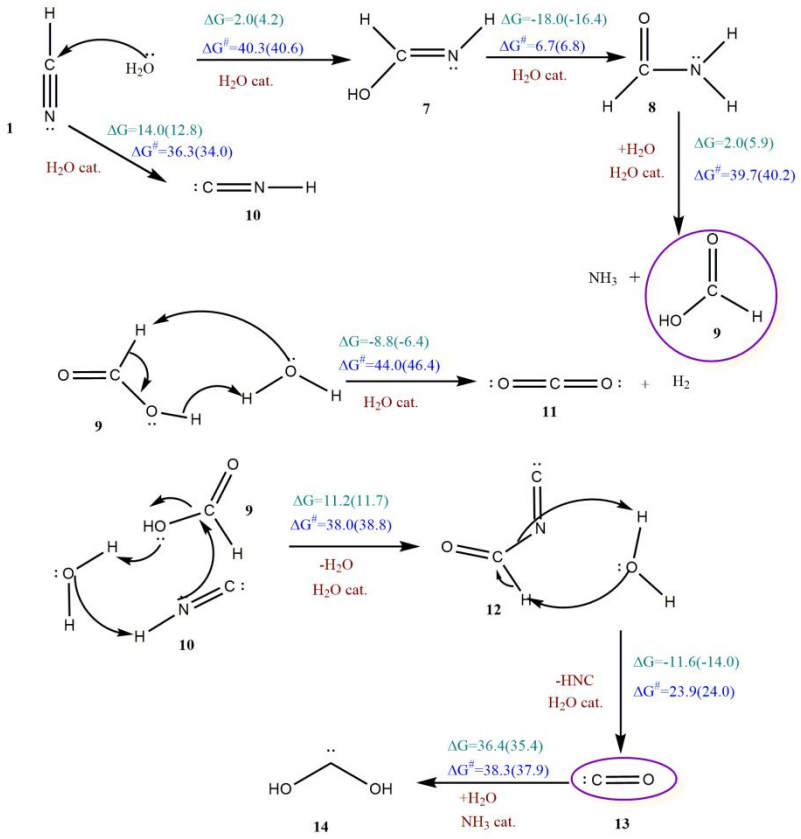


Figure S4. The sequence of elementary reaction steps derived from the AINR: the formation of HCOOH, CO_2 and CO starting from HCN and H_2O. Molecules labelled 'cat.', shown in brown, participate catalytically as proton shuttles. Values (in kcal/mol) calculated at the B3LYP-D2/6-311++g(d,p)+PCM (ε=80.0) and the M06-2X/6-311++g(d,p)+PCM(ε=80.0) (values shown in parenthesis) levels of theory with the Gaussian09 software package.[6]

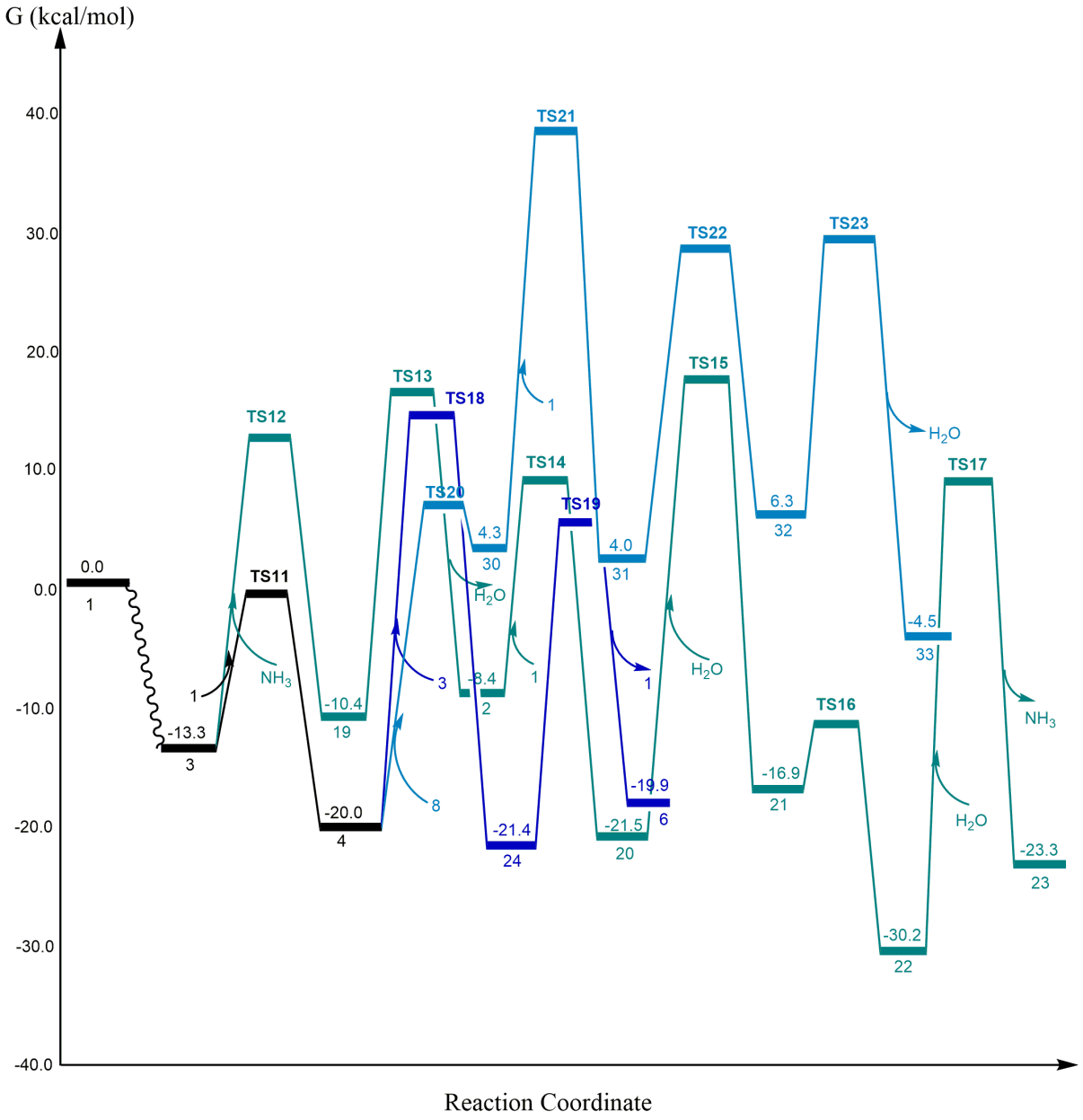


Figure S5. The reaction free energy profile diagram for the formation of the RNA precursors: glycoaldehyde and oxazole; and protein precursors: the glycine molecules *via* intermediate species formaldehyde, formaldimine and glycolonitrile beginning from HCN and H_2O. The relative free energies of the reactants and products for each elementary step are represented with respect to the beginning reactants and the barrier has been calculated from the reactant species of each elementary step. The values (in kcal/mol) have been obtained at the B3LYP-D2/6-311++g(d,p)+PCM (Ɛ=80.0) level of theory by Gaussian09 software package.[6]

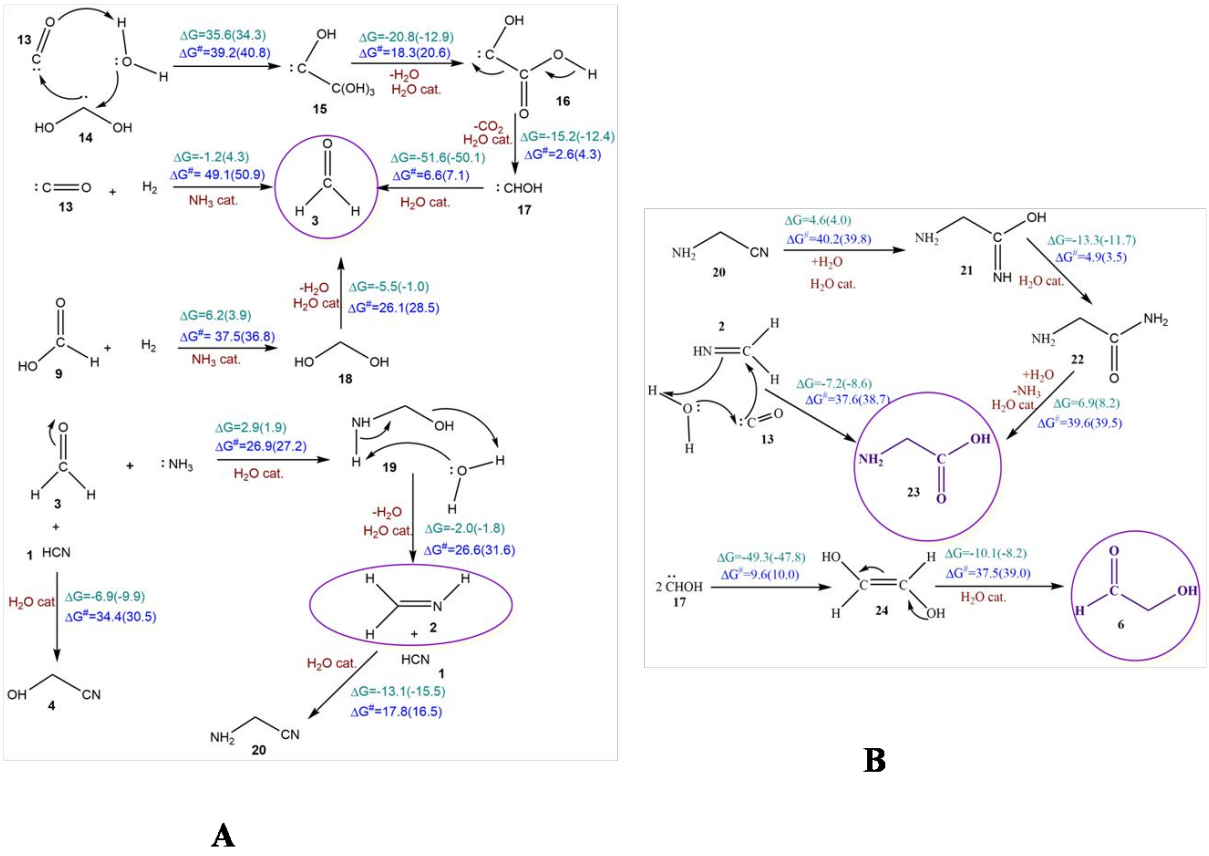


Figure S6. (**A**) The sequence of elementary reaction steps derived from the AINR: the formation of formaldehyde, formaldimine, glycolonitrile and aminoacetonitrile. (**B**)The formation of the target species: glycine and sugar. Values (in kcal/mol) calculated at the B3LYP-D2/6-311++g(d,p)+PCM (ε=80.0) and the M06-2X/6-311++g(d,p)+PCM(ε=80.0) (values shown in parenthesis) levels of theory with the Gaussian09 software package.[6]

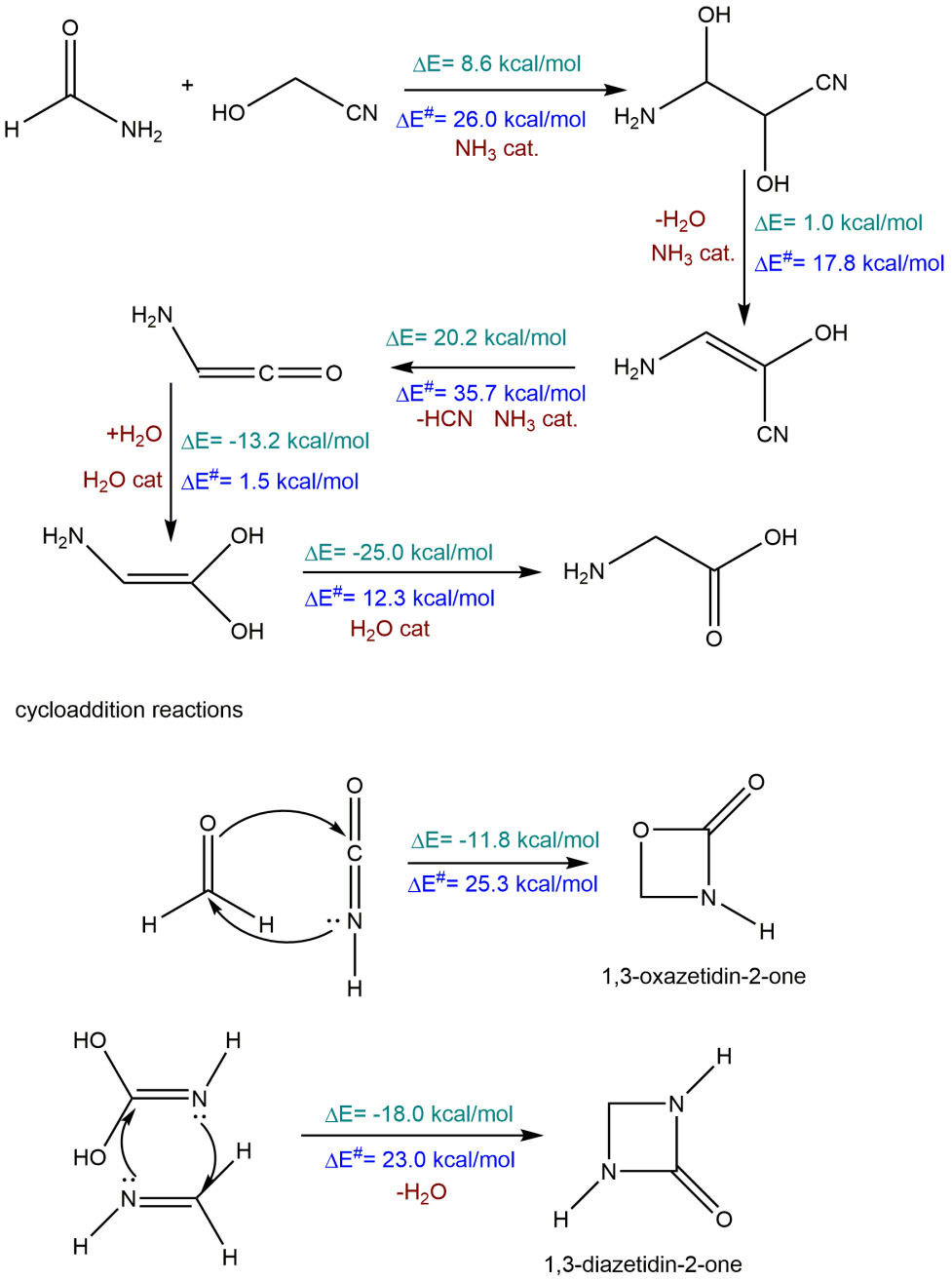
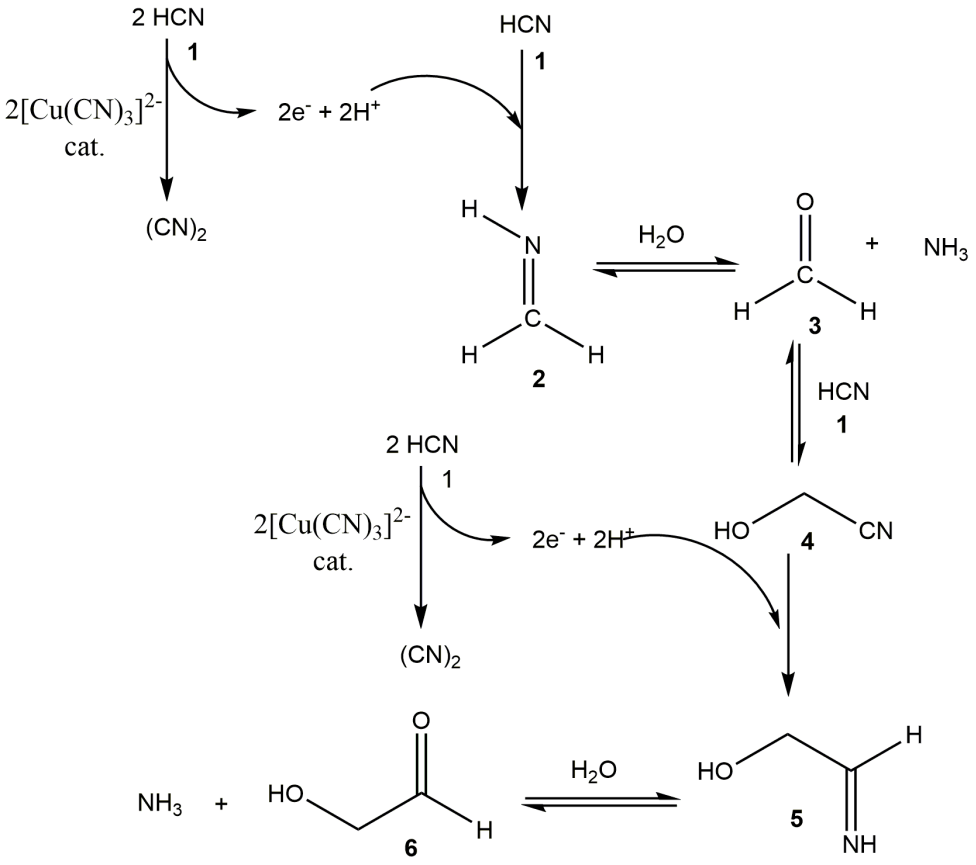
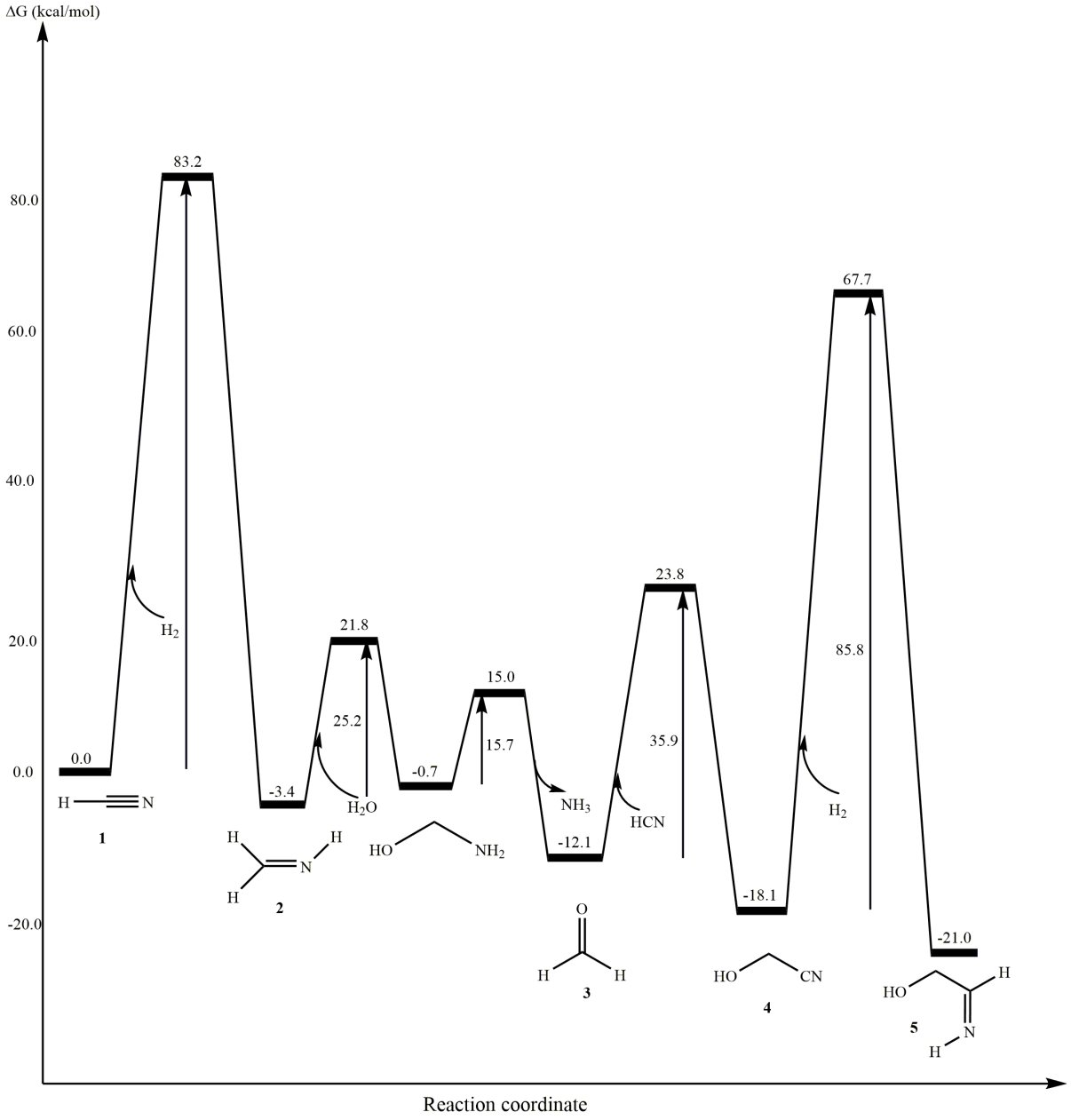


Figure S7. Alternative pathways for the formation of glycine and cycloaddition reactions observed during the *ab initio* nanoreactor simulations. Reaction energies (ΔE) are shown in green and the activation barriers ($\Delta E^{\#}$) are shown in blue, calculated at the B3LYP-D3/TZVP+COSMO (ε=80.0) level of theory with DFT by the Turbomole 7.0 software package.[1]


2 HCN
1
2[Cu(CN)3]2-
cat.
(CN)2
2e- + 2H+
HCN
1
H
N
C
H
H
2
H2O
O
C
H
H
3
+
NH3
HCN
1
2 HCN
1
2[Cu(CN)3]2-
cat.
(CN)2
2e- + 2H+
HO
CN
4
NH3
+
HO
O
H
6
H2O
HO
H
5
NH

(A)



(B)

Figure S8. (A) The sugar synthesis pathway proposed by Sutherland and co-workers[26] from HCN and water. It involves a photoredox cycling of the copper cyanide complex catalyst, producing two protons and two hydrated electrons from HCN, which further reduce another HCN molecule **1** to aldimine **2** in one step and glycolonitrile **4** to imine **5** in another step.**(B)**The reaction free energy profile diagram for the sugar formation *via* the pathway proposed by Sutherland and co-workers.[26] The relative free energy of the reactants and the products for each elementary step are represented with respect to the beginning reactants and the barrier has been calculated from the reactant species, for each elementary step. The values (in kcal/mol) have

been obtained at the B3LYP-D3/TZVP+COSMO(Ɛ=80.0)//RI-CC2/TZVP+COSMO(Ɛ=80.0) level of theory by the use of the Turbomole 7.0 software package.[1]

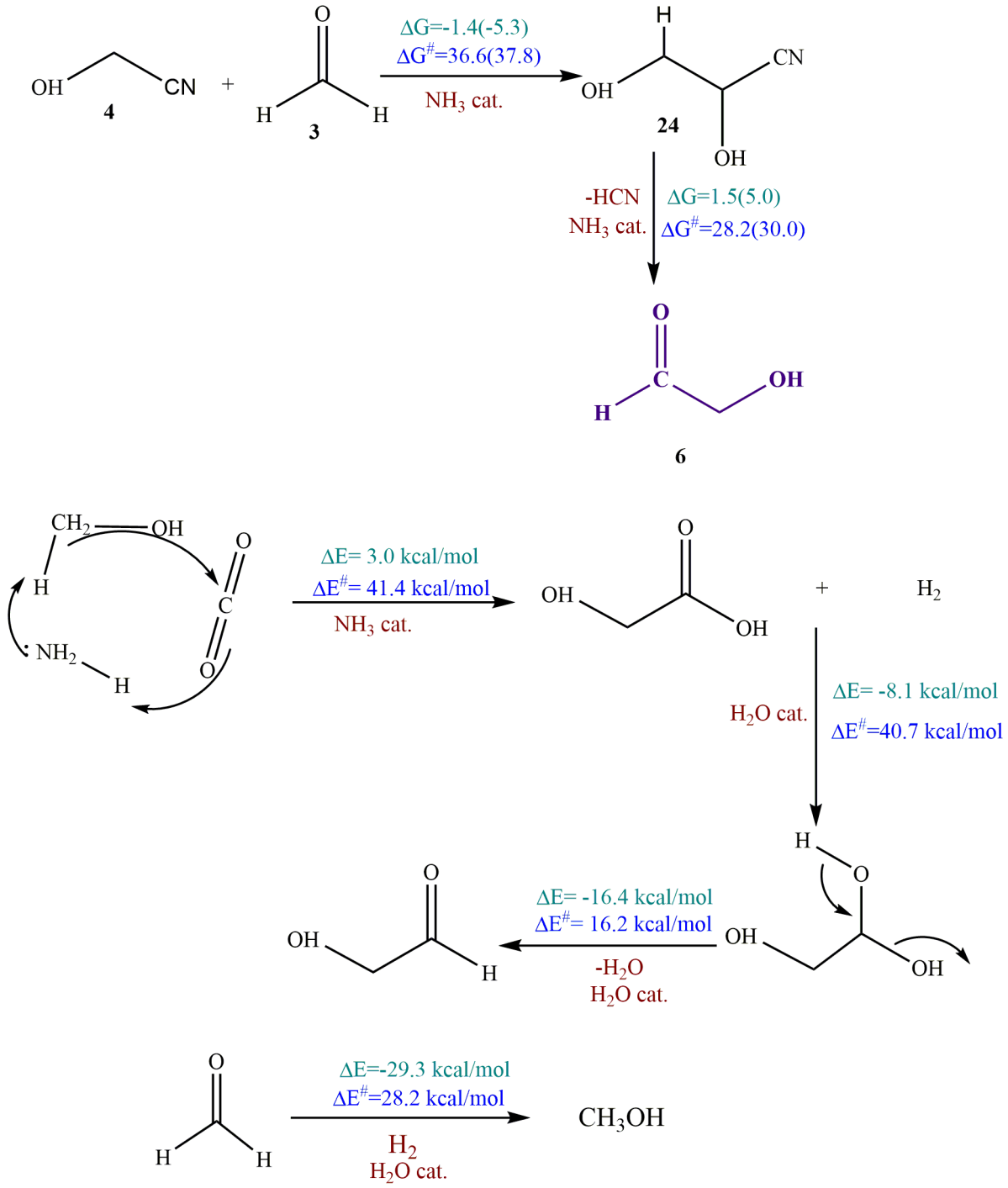


Figure S9. Alternative pathways for the formation of sugar during the *ab initio* nanoreactor dynamics. The reaction energies (ΔE) are shown in green and the activation barriers ($\Delta E^{\#}$) are shown in blue, calculated at the B3LYP-D3/TZVP+COSMO (ε=80.0) level of theory with DFT by the Turbomole 7.0 software package.[1]

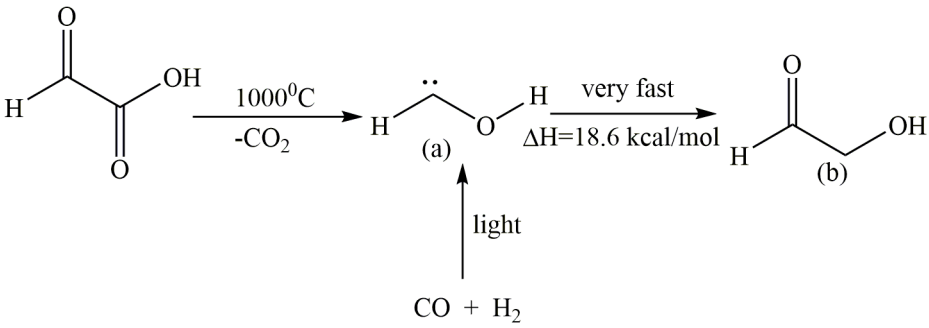


Figure S10. Schreiner and co-workers have proposed[27] a new reaction pathway for sugar formation *via* hydroxyl methylene in the gas phase or on surfaces in the absence of a base.

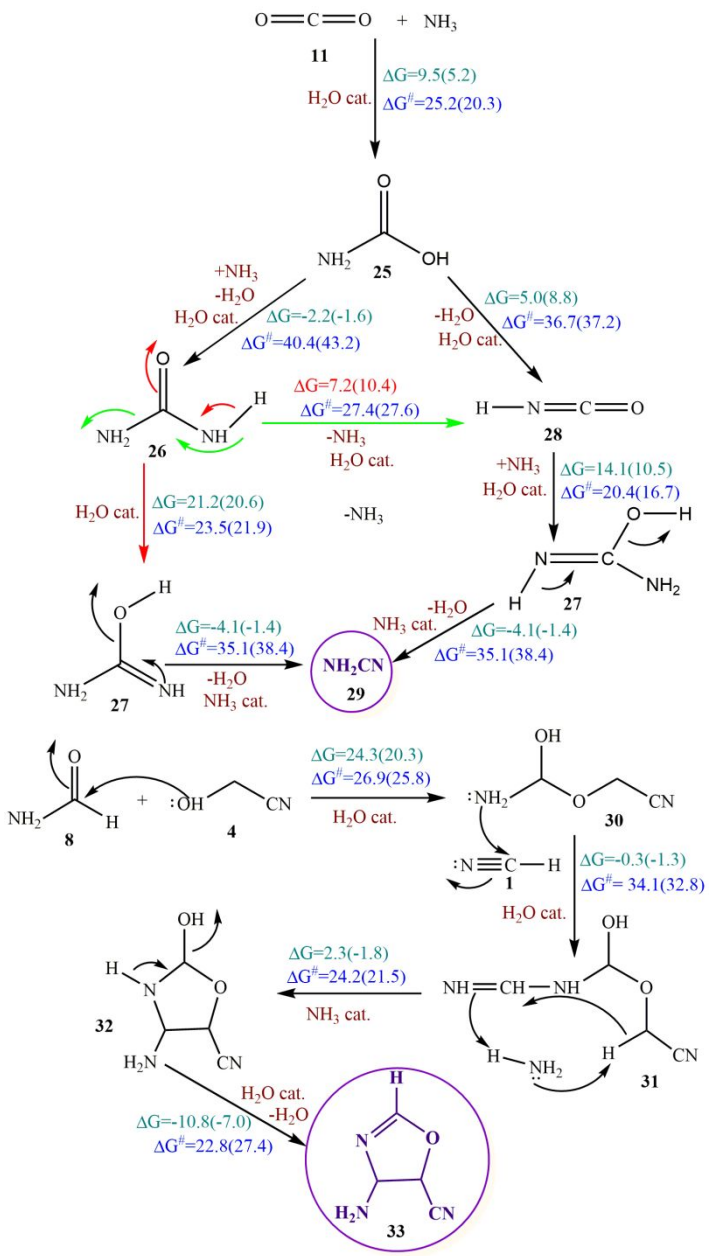


Figure 11. The formation of the target species: cyanamide and the oxazole derivative. Values (in kcal/mol) have been calculated at the B3LYP-D2/6-311++g(d,p)+PCM (ε=80.0) and the M06-2X/6-311++g(d,p)+PCM(ε=80.0) (values shown in parenthesis) levels of theory with the Gaussian09 software package.[6]

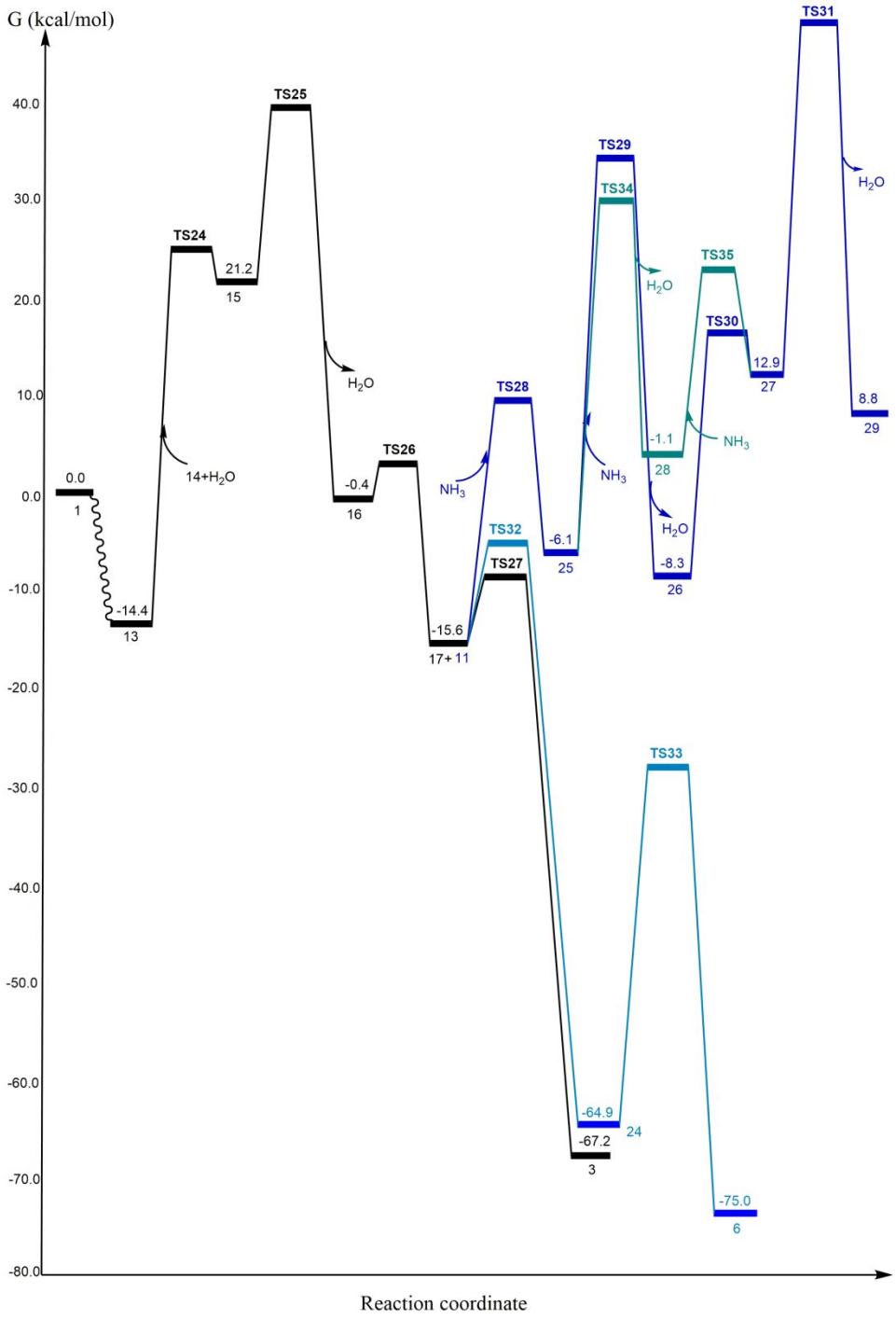


Figure S12. The reaction free energy profile diagram for the formation of the RNA precursors: cyanamide and sugar, starting from HCN and H_2O and with CO_2, urea, formaldehyde,

glycolonitrile and other intermediates formed along the route. The relative free energy values of the reactants and products for each elementary step are represented with respect to the beginning reactants and the barrier has been calculated from the reactant species for each elementary step reaction. The values (in kcal/mol) have been represented at the B3LYP-D2/6-311++g(d,p)+PCM(Ɛ=80.0) level of theory with DFT calculated with the Gaussian09 software package.[6]

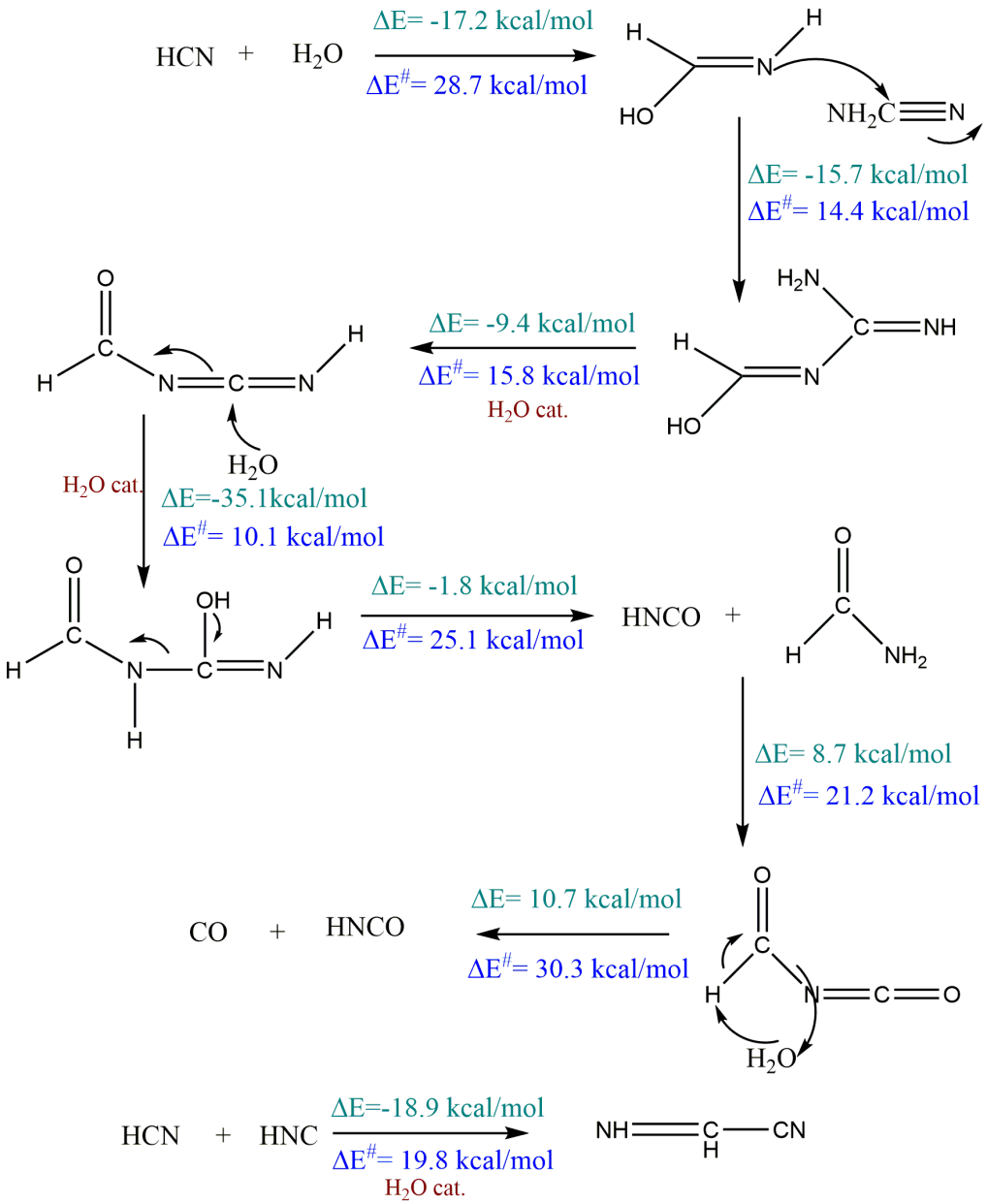
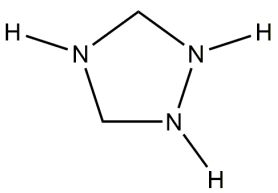
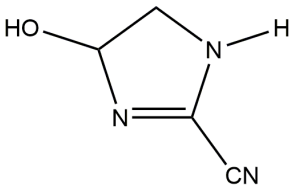


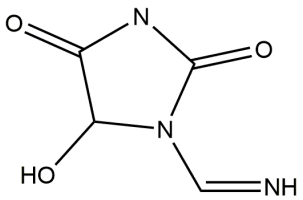
Figure S13. Alternative pathways for the formation of CO, iminoacetonitrile and isocyanic acid during the nanoreactor dynamics. Reaction energies (ΔE) are shown in green and the activation barriers ($\Delta E^{\#}$) are shown in blue, calculated at the B3LYP-D3/TZVP+COSMO (ε=80.0) level of theory with DFT with the Turbomole 7.0 software package.[1]



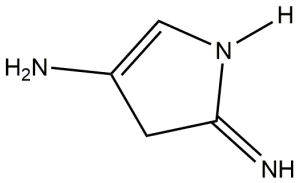
1,2,4-triazolidine



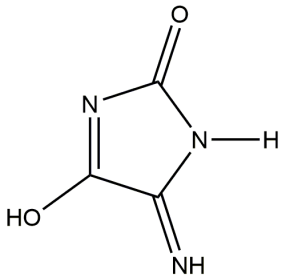
4-hydroxy-4,5-dihydro-1*H*-imidazole-2-carbonitrile



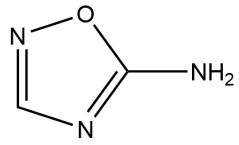
5-hydroxy-1-(iminomethyl)-3λ^2-imidazolidine-2,4-dione



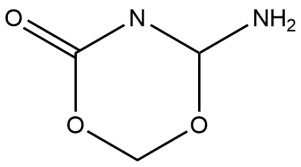
2-imino-4,5-dihydro-1*H*-pyrrol-3-amine



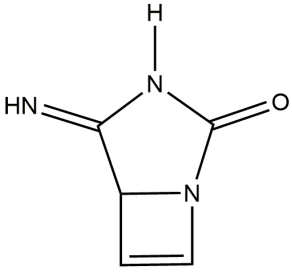
4-hydroxy-5-imino-1,5-dihydro-2*H*-imidazol-2-one



1,2,4-oxadiazol-5-amine



6-amino-1,3,5λ^2-dioxazinan-4-one



4-imino-1,3-diazabicyclo[3.2.0]hept-6-en-2-one

Figure S14. A selection of heterocyclic products that were discovered from the *ab initio* nanoreactor simulations, including the ribonucleotide precursor oxazole discussed in the main manuscript. Here, we have shown that apart from the oxazole, a lot of diverse cyclic organic compounds were also formed during the simulations.

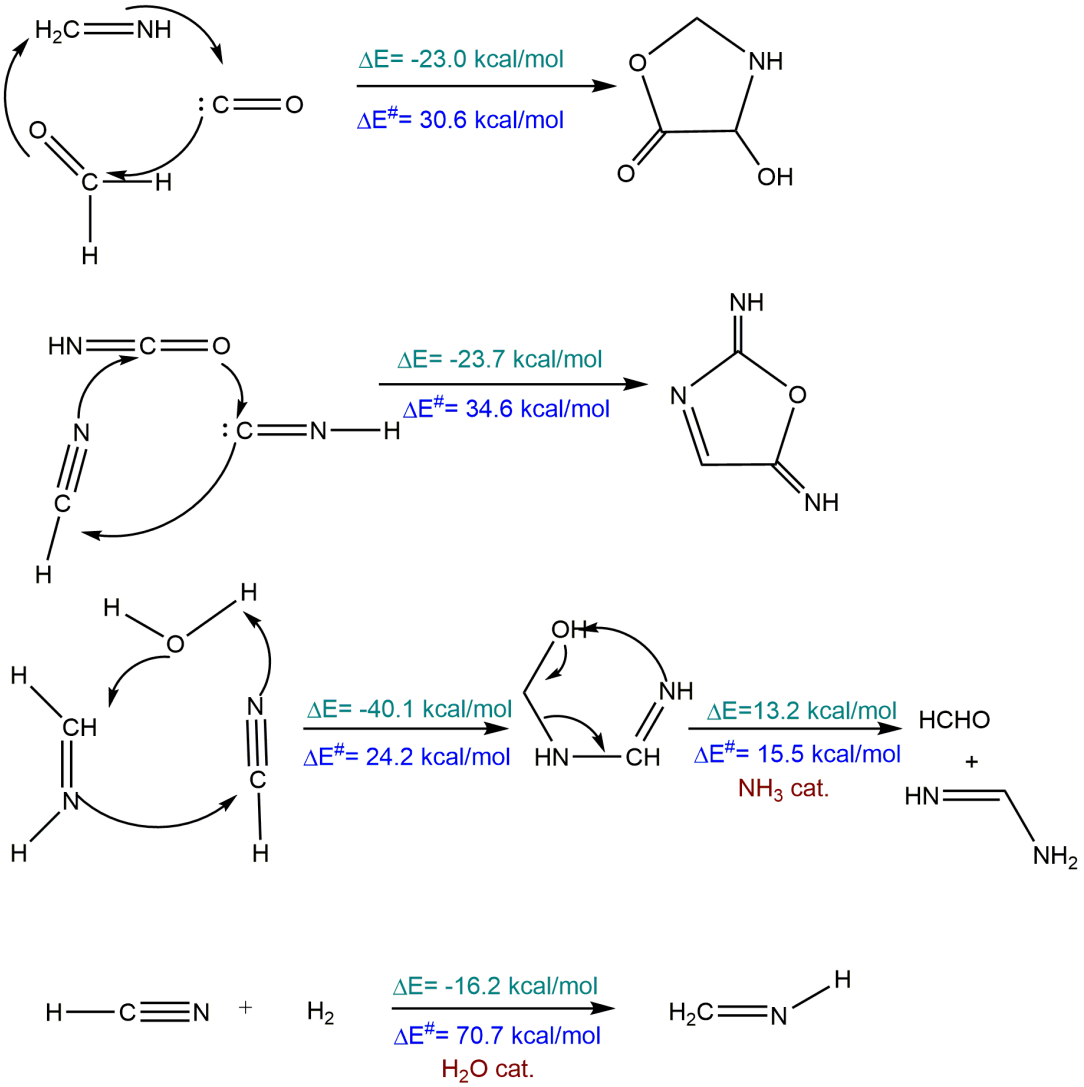


Figure S15. The formation of some cyclic species during the nanoreactor dynamics. The reaction energies (ΔE) are shown in green and the activation barriers ($\Delta E^{\#}$) are shown in blue, calculated at the B3LYP-D3/TZVP+COSMO(ε=80.0) level of theory with DFT with the Turbomole 7.0 software package.[1]

The Cartesian (x, y, z) Coordinates of all the Transition State Structures Obtained at the B3LYP/6-311++g(d,p) Level of Theory Employing Gaussian09.

TS1

C	-1.583501	1.695475	-0.070093
H	-1.496212	1.032458	-0.918951
N	-1.908574	2.803627	0.245547
O	-0.940711	0.667461	1.208983
H	-1.432288	-0.165316	1.230331
H	-1.185021	1.466773	2.169689
O	-1.531747	2.455193	2.660794
H	-0.793850	2.926461	3.070441

H	-1.756233	2.875793	1.601047

TS2

C	-3.772680	1.096353	0.161883
N	-3.472497	-0.142836	-0.069751
O	-2.060548	-0.272727	1.898929
O	-3.277984	1.724570	1.186281
H	-4.443183	1.664124	-0.480198
H	-2.689624	-0.515073	0.796074
H	-3.880166	-0.583268	-0.883495
H	-1.101692	-0.305459	1.805648
H	-2.629861	0.995003	1.682427

TS3

C	-3.157641	2.089198	0.077010
O	-2.624658	1.862161	2.128105
N	-2.975187	0.607060	-0.165624
O	-4.262220	2.562966	0.014767
O	-1.151581	0.061208	1.580007
H	-2.741203	0.454227	-1.149351
H	-2.213148	2.617231	-0.067974
H	-2.169996	0.202312	0.478843
H	-1.266486	-0.722848	2.127189
H	-3.857192	0.129257	0.028687
H	-2.222895	2.704524	2.368306
H	-1.779697	0.957068	2.012076

TS4

C	-2.093137	-0.178822	0.733696
H	-1.522050	-1.038656	0.399714
O	-1.353859	1.044285	-0.065287
H	-1.843265	1.864631	0.114788
O	-3.245324	-0.009473	0.924027
C	-0.508858	-0.016902	3.412713
N	-1.029829	0.040158	2.363883
O	1.037855	1.338025	0.691408
H	1.739159	0.967104	0.140726
H	-0.358122	1.170918	0.207980
H	1.085359	0.891840	1.550819

TS5

C	0.265883	-0.060778	0.109484
H	-0.318997	-0.111000	1.255285
N	0.916861	1.923894	0.825071
C	0.606440	2.546206	1.770861
O	0.644564	-0.291245	-0.926096
O	-0.798482	-0.027510	2.451016
H	-1.748347	-0.206217	2.518908
H	-0.640721	0.890702	2.738862

TS6

C	-2.859470	1.160388	-0.129049
O	-2.561193	1.702987	0.926984
O	-3.804403	0.288736	2.582354
H	-3.219061	1.053897	2.189267
H	-4.591193	0.649023	3.023806
O	-3.744075	-0.062254	0.209116
H	-4.045255	-0.130780	1.645161
H	-4.052282	-0.488519	-0.597789

TS7

C	-5.809912	0.417112	0.326298
N	-5.669334	0.815829	1.416629
O	-4.795083	3.097678	0.088970
H	-5.054885	2.280639	-0.411407
H	-4.819449	2.811552	1.033213
H	-3.890265	3.375982	-0.157171

TS8

C	-1.883229	-0.125875	-0.082735
O	-2.004915	-1.126618	0.576665
O	-1.414924	0.308096	-1.112714
O	-2.303296	2.835317	-0.319381
H	-2.530842	0.855297	0.631860
H	-1.929143	2.537980	-1.167924
H	-3.135312	3.301578	-0.501173
H	-2.496015	1.778178	0.285364

TS9

C	-3.741809	1.587891	-0.283426
O	-3.273442	0.445349	-0.417794
O	-4.863023	1.717311	0.552249
N	-1.243309	1.081860	1.515306
H	-3.774196	2.304395	-1.118993
H	-5.147830	2.640826	0.544417
H	-2.034451	1.963752	0.997612
H	-1.556865	0.196232	1.122701
H	-2.742151	2.485933	0.500009
H	-0.274793	1.250152	1.253966
H	-1.317398	1.057544	2.529465

TS10

C	-3.216892	0.578888	0.120344
H	-2.879388	-0.365751	-0.319046
H	-3.163681	1.417088	-0.586908
O	-2.851642	0.811457	1.357371
H	-3.840667	1.547865	1.845261
O	-4.872873	0.396833	0.059621

H	-5.087384	-0.507453	0.330073
O	-4.925404	1.847900	1.947840
H	-5.143690	1.138594	0.944898
H	-5.045588	2.793062	1.790994

TS11

C	0.142995	0.153504	-0.235388
O	0.172596	1.001883	-1.141184
C	0.872558	1.529873	1.717037
N	1.793770	2.198359	1.440427
O	2.349767	2.007034	-1.143073
H	1.050489	-0.353262	0.093684
H	-0.821146	-0.184905	0.144972
H	2.351914	2.256178	-0.162431
H	1.278786	1.484117	-1.280287
H	2.449475	2.808811	-1.676225

TS12

C	-2.510746	0.632792	0.008406
H	-2.908200	1.209313	0.855228
H	-3.184716	-0.203592	-0.223294
O	-2.086881	1.380511	-1.079812
N	-1.140568	0.079379	0.382108
H	-1.033359	-0.928487	0.283089
H	-0.776810	0.382432	1.283732
H	-0.953888	0.790555	-0.597503

TS13

N	-2.738241	0.983006	0.011462
H	-3.667342	0.490003	0.090604
O	-4.693495	-0.105430	1.329600
H	-3.894748	0.507461	2.221861
O	-3.121981	1.085554	2.783719
C	-2.608250	1.976788	0.832202
H	-3.488488	2.545281	1.090025
H	-1.627717	2.377708	1.055721
H	-2.413337	0.474831	3.017032
H	-1.918993	0.452887	-0.260529
H	-5.623207	0.130426	1.405886

TS14

C	-0.490374	2.148493	0.865879
N	-1.392566	1.730071	0.057615
C	-0.912327	1.052353	3.199553
N	-1.020488	1.524532	4.263478
O	-0.965992	-0.861079	0.908647
H	-2.225242	2.274070	-0.138679
H	-0.560940	3.142656	1.287179
H	0.384413	1.538518	1.027784

H	-0.913164	-0.350411	1.748682
H	-0.273535	-1.528165	0.957485
H	-1.345120	0.759688	-0.261441

TS15

N	-4.257932	-3.420722	1.682109
H	-4.146400	-3.987438	0.847559
H	-5.249358	-3.403497	1.901014
C	-3.777080	-2.067693	1.402812
H	-4.387604	-1.500165	0.686836
H	-2.767460	-2.116240	0.988982
C	-3.707049	-1.274048	2.655274
N	-3.893195	-1.247797	3.832866
O	-2.984094	0.335397	2.004158
H	-3.583263	0.715361	1.348319
H	-3.056472	0.923883	3.107921
O	-3.160559	1.054754	4.269146
H	-3.602166	-0.121977	4.292368
H	-3.847770	1.696162	4.489392

TS16

N	-1.810612	0.773186	0.002091
H	-1.471534	1.045083	-0.914226
H	-2.788036	1.044359	0.049130
C	-1.041329	1.474228	1.030217
H	-1.251020	2.549302	1.096819
H	0.023544	1.359204	0.802712
C	-1.248298	0.873616	2.400820
O	-1.123571	-0.419874	2.524244
H	-1.260501	-0.630043	3.565326
N	-1.506941	1.590354	3.454309
H	-1.564807	0.867413	4.428288
O	-1.571348	-0.324161	4.952753
H	-0.844287	-0.468883	5.568280
H	-1.606761	2.589417	3.333361

TS17

N	-3.240530	-0.652573	-0.151335
H	-2.686396	-1.503589	-0.165488
H	-2.913662	-0.069628	-0.914701
C	-3.036723	0.049080	1.108624
H	-3.601716	0.980633	1.122708
H	-1.979883	0.304022	1.308705
C	-3.456182	-0.803900	2.281716
N	-3.240077	-0.051691	3.597714
H	-3.569711	-0.643196	4.361700
H	-2.241579	0.120873	3.737628
O	-3.488852	-2.006186	2.326082
O	-5.551048	-0.121998	2.228582
H	-6.008715	-0.735118	2.814703

H	-3.791377	0.883231	3.600485
O	-4.885677	1.956718	3.249861
H	-5.364187	1.039778	2.789730
H	-5.412604	2.248271	4.000839

TS18

C	-0.899806	2.106904	1.039327
O	-0.389894	3.232872	1.093292
C	-0.562448	0.752017	-1.003034
O	-1.437516	-0.201898	-0.390067
C	-1.185605	1.656917	-1.894208
N	-1.674958	2.487290	-2.555352
N	1.895876	2.284731	-0.051948
H	-1.732909	-0.835646	-1.055043
H	1.097456	1.664071	-0.352472
H	0.311324	0.269003	-1.451235
H	1.459728	2.958264	0.598814
H	-1.939531	1.964012	0.718201
H	-0.425424	1.233311	1.510523
H	2.618678	1.739091	0.416462
H	2.297135	2.769364	-0.854249

TS19

O	-4.685296	1.784782	0.033133
H	-4.877757	2.695995	0.275353
C	-3.273385	1.614593	0.042818
H	-2.780550	2.452444	-0.471181
H	-2.869867	1.548582	1.058953
C	-2.884931	0.380863	-0.753056
H	-3.553273	0.164117	-1.596680
C	-3.845755	-1.075228	0.503030
N	-3.737040	-1.898597	1.325343
O	-1.684878	0.021439	-0.794608
H	-1.001807	-0.645714	0.548234
H	-1.700144	-1.699690	1.640323
N	-0.801083	-1.259988	1.388589
H	-0.113873	-1.975187	1.151947
H	-0.454845	-0.703670	2.169559

TS20

C	-3.176469	1.297337	-2.004273
O	-1.941098	1.866310	-1.787905
C	-2.326302	4.135531	-1.297932
O	-2.625010	3.983171	-0.048825
C	-3.789516	0.765544	-0.752473
N	-4.210254	0.366737	0.247275
N	-3.242297	4.572000	-2.125600
O	-0.998111	2.147798	0.507234
H	-1.332919	1.826779	-0.423929
H	-3.149892	0.450624	-2.711883

H	-3.006720	4.770041	-3.086845
H	-4.205186	4.667390	-1.829057
H	-1.302367	4.086315	-1.640075
H	-1.285371	1.501039	1.161115
H	-1.932028	3.348483	0.378543
H	-3.911324	2.015459	-2.419875

TS21

N	-3.601618	1.984418	0.527409
H	-3.105243	1.177894	0.142969
H	-5.541564	2.881540	-0.546995
C	-3.019760	2.335153	1.839409
H	-2.007837	2.728504	1.693110
O	-2.994284	1.126894	2.545508
C	-2.322067	1.236710	3.801763
H	-2.847904	1.929286	4.466431
H	-1.291602	1.586735	3.664741
C	-2.298115	-0.091743	4.412456
N	-2.270964	-1.130814	4.908498
O	-3.846596	3.263278	2.456520
H	-3.371258	4.088309	2.607027
C	-3.555065	3.187557	-0.519591
H	-2.516682	3.504672	-0.631923
N	-4.595981	3.586694	-1.034967
O	-6.159742	1.997935	0.072048
H	-6.750384	2.396287	0.720741
H	-4.661253	1.768585	0.599395

TS22

C	0.965030	0.678380	0.616172
C	3.645185	0.141573	0.364800
O	3.380722	-1.042877	1.135468
C	2.122777	-1.094567	1.783060
N	1.029559	-0.637567	0.970502
C	4.275732	1.180692	1.035366
N	4.810230	2.061525	1.603779
O	2.119409	-0.301440	2.951342
N	0.355843	1.061872	-0.472386
N	1.664801	0.105808	-2.455185
H	4.014765	-0.103487	-0.622911
H	1.968975	-2.150627	2.015582
H	2.978324	-0.420939	3.374180
H	0.719308	-1.307547	0.277084
H	1.318092	1.369538	1.371951
H	0.338197	2.074433	-0.543800
H	1.595374	-0.884792	-2.681003
H	1.550879	0.651652	-3.306798
H	0.897552	0.461078	-1.596618
H	2.585912	0.301815	-2.053809

TS23

C	-3.468111	0.486486	0.058117
C	-2.747773	1.163765	-1.158806
C	-2.671919	2.553145	0.600869
H	-3.417891	1.289893	-2.007365
H	-4.545783	0.550933	-0.091663
O	-2.369689	2.488517	-0.683025
N	-3.088928	1.438332	1.137557
N	-3.109909	-0.879560	0.252509
H	-2.163104	-0.985780	0.602789
H	-3.753141	-1.345501	0.881120
C	-1.531579	0.470570	-1.593429
N	-0.577272	-0.072762	-1.939023
H	-3.726285	1.564127	1.985722
H	-2.337804	3.418973	1.147079
O	-4.680750	3.612875	0.656766
H	-4.572664	4.564160	0.764301
O	-4.992564	2.323563	2.672301
H	-4.787363	2.792759	3.487313
H	-4.914783	3.069961	1.741280

TS24

C	0.028026	0.249057	-0.031267
O	0.253185	0.896474	-1.081539
C	1.396363	0.081789	0.751444
O	1.831984	-1.184345	0.778653
O	2.434111	0.943253	-0.255378
H	3.065847	0.328158	-0.656723
H	1.555708	1.095951	-0.952168
H	2.433024	-1.321775	1.527826
O	1.475826	0.680516	1.952357
H	0.987186	1.516658	1.957457

TS25

C	1.042142	0.437315	-0.036355
O	-0.675234	1.989952	-3.095397
H	0.060983	1.114102	-3.075365
H	-1.581170	1.706142	-2.909098
O	0.014826	-0.300027	-0.347290
H	-0.328855	-0.728965	0.450744
C	1.463995	1.254142	-1.250202
O	2.843685	1.342660	-1.331737
H	3.048053	2.217559	-1.692864
O	0.806862	2.386950	-1.219555
H	-0.208250	2.370424	-2.201844
O	1.053295	0.433337	-2.605637
H	1.838839	0.427686	-3.169206

TS26

C	-0.965026	0.678081	-0.066795

O	-0.815810	1.331409	-1.124591
C	-0.104244	1.508278	1.565309
O	-0.599532	0.847516	2.732778
O	0.562922	2.490648	1.456699
O	-1.474241	-0.704855	1.581215
H	-1.109121	-0.267680	2.606679
H	-1.193278	0.890287	-1.917162
H	-1.240907	-0.521413	0.299133
H	-1.963394	-1.530939	1.714812

TS27

C	0.537632	0.616039	-0.094191
H	0.868420	-0.421580	0.040406
O	1.036790	1.146091	-1.136474
H	0.612887	2.100092	-1.084491
O	-0.315734	2.894620	-0.050844
H	-0.182478	1.698813	0.317409
H	-1.212487	3.046970	-0.367613

TS28

C	-3.674470	0.123442	0.105866
O	-3.866467	1.302444	0.563355
O	-4.417085	-0.557819	-0.576964
N	-2.325163	-0.431086	0.490896
H	-2.450322	-1.068693	1.277903
H	-1.635392	0.686101	0.921173
H	-1.958318	-0.984802	-0.280195
O	-1.639649	1.771139	1.316600
H	-1.117536	2.358129	0.753691
H	-2.746035	1.776556	1.015516

TS29

N	-3.431986	0.073230	0.131731
H	-4.393477	0.282194	-0.105092
H	-2.805402	0.250076	-0.643877
C	-2.986186	0.657680	1.297477
O	-4.075421	2.467132	1.232727
O	-1.829285	0.898536	1.569866
N	-3.906895	0.300927	2.468338
H	-4.888526	0.285860	2.189934
H	-3.646167	-0.614808	2.840684
O	-3.679774	2.744050	3.704775
H	-3.840724	2.844904	2.689144
H	-4.460064	3.092264	4.148299
H	-3.792342	1.049201	3.195420
H	-3.549565	2.966797	0.599595

TS30

N	-3.080825	0.879925	0.156112

H	-2.743893	-0.065892	0.074370
H	-2.811547	1.549083	-0.548140
C	-3.610390	1.320244	1.309457
N	-3.884902	0.574715	2.353137
H	-3.678796	-0.412928	2.307634
H	-4.394227	1.278563	3.216543
O	-3.883881	2.609725	1.359377
O	-4.757888	2.411717	3.694918
H	-5.711647	2.457187	3.822584
H	-4.313417	2.776204	2.302794

TS31

N	-2.800724	1.127217	0.007459
H	-2.665600	0.191382	-0.355678
H	-3.130337	1.777463	-0.692713
C	-3.515960	1.140566	1.189559
N	-3.839948	0.419775	2.092209
H	-4.529561	1.438311	3.366080
O	-3.968416	2.861906	1.245309
H	-4.506366	2.933249	2.778622
H	-3.186671	3.383864	1.024765
N	-4.795245	2.422465	3.671676
H	-4.259928	2.725856	4.482554
H	-5.791608	2.507704	3.861755

TS32

C	0.743907	0.622179	-0.959804
O	0.869227	0.652609	-2.256473
C	-1.980047	-0.722136	-0.222105
O	-1.726438	-0.478333	1.034694
H	1.767297	0.865683	-0.599723
H	0.019931	0.449872	-2.683007
H	-1.900091	0.233058	-0.780629
H	-1.463300	0.444452	1.221109

TS33

C	-1.961260	0.162265	0.000026
H	-1.534674	0.443671	-0.975567
C	-1.316985	0.660981	1.128706
H	-1.382603	0.123723	2.075158
O	-3.087312	-0.450394	0.019765
O	-0.167427	1.461627	0.956328
H	0.619895	0.909357	1.052021
O	-3.828594	1.270494	1.660844
H	-3.854365	0.404317	1.087002
H	-2.704051	1.355495	1.545708
H	-4.044663	1.058986	2.583149

TS34

C	-1.625057	0.118459	1.131156
O	-1.354673	-0.930770	2.250885
N	-2.192745	-0.516254	0.153709
O	-1.267769	1.238290	1.432229
O	-1.784975	-2.886505	0.974264
H	-2.060395	-2.114069	0.332698
H	-2.428694	0.113136	-0.605463
H	-0.560954	-0.684587	2.747820
H	-1.056292	-3.383339	0.578575
H	-1.417070	-2.012063	1.795628

TS34_1

C	-3.339312	2.112245	-0.020311
O	-2.532842	2.138157	-0.934397
N	-4.608614	1.784911	0.065753
H	-5.041030	1.487591	-0.799802
H	-5.090588	1.871825	1.174894
N	-2.836351	2.545752	1.368127
H	-2.041345	1.963351	1.634229
H	-2.506477	3.510725	1.318851
O	-5.159881	2.171353	2.434152
H	-3.667717	2.443682	2.077516
H	-5.289801	1.390151	2.982188

TS35

C	-0.691486	2.174675	0.471656
N	-1.385754	1.739186	-0.502269
O	0.481336	2.737983	0.487930
N	-1.282088	2.037962	1.839314
H	0.707388	3.121982	1.580934
H	-2.252309	2.348280	1.827607
H	-1.284571	1.054812	2.111281
H	-0.884974	1.898653	-1.376569
O	0.577099	3.300815	2.811619
H	-0.515875	2.700813	2.568670
H	0.450820	4.230241	3.036646

References

1. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system Turbomole. *Chem. Phys. Lett*. **1989**, *162*, 165−169.

2. Ansgar, S.; Christian, H.; Reinhart, A. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr *J. Chem. Phys*.**1994**, *100*, 5829−5835.

3. Becke, A. D. Density functional thermochemistry. III. The role of exact exchange *J. Chem. Phys.* **1993**, *98*, 5648–5652.

4. Klamt, A.; Schuurmann, G. J. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient *Chem. Soc., Perkin Trans*. **1993**, *2*, 799−805.

5. Grotendorst, J.; Bl¨ugel, S.; Marx, D Beyond Hartree-Fock: MP2 and coupled-luster methods for large systems *Computational Nanoscience,* **2006,** 31, 245-278.

6. Gaussian 09, Revision **E.01**; Frisch , M. J.; Trucks, G. W.; Schlegel, H. B.;Scuseria,G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.;Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.;Jr.,Peralta, J. E.; Ogliaro, F.,Bearpark, M.; Heyd, J. J. E.; Brothers,Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Raghavachari, J.K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., *Wallingford CT*, **2009.**

7. McLean, A. D.; Chandler, G. S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, *Z*=11–18 *J. Chem. Phys.* **1980**, *72*, 5639-5648.

8. Hepburn, J.; Scoles, G.; Penco, R. A simple but reliable method for the prediction of intermolecular potentials *Chem. Phys. Lett*. **1975**, *36*, 451–456.

9. Ahlrichs, R.; Penco, R.; Scoles, G. Intermolecular forces in simple systems *Chem. Phys*.**1977**, *19*, 119–130.

10. Grimme, S. Accurate description of Van der Waals complexes by density functional theory including empirical corrections *J. Comput. Chem*. **2004**, *25*, 1463–1473.

11. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem*. **2006**, *27*, 1787–1799.

12. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu *J. Chem. Phys*. **2010**, *132*, 154104-154119.

13. Zhao, Y.; Truhlar, D. G. Exploring the limit of accuracy of the global hybrid meta density functional for main-group thermochemistry, kinetics, and noncovalent interactions. *Theor. Chem. Acc.***2008**, *120*, 215–241.

14. Zhao, Y.; Truhlar, D. G. Density functional for spectroscopy: no long-range self-interaction error, good performance for Rydberg and charge-transfer states, and better performance on average than B3LYP for ground states**.** *J. Phys. Chem.* A **2006**, *110,* 13126–13130.

15. Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125*, 194101.

16. Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models *Chem. Rev.***2005**, *105*, 2999-3094.

17. Mammen, M.; Shakhnovich, E. I.; Deutch, J. M.; Whitesides, G. M. Estimating the entropic cost of self-assembly of multi particle hydrogen-bonded aggregates based on the cyanuric acid• melamine lattice *J. Org. Chem.***1998**, *63*, 3821−3830.

18. Hagberg, A. A.; Schult, D. A.; Swart, P. J. Exploring network structure, dynamics, and function using Networkx *Proceedings of the 7th Python in Science Conference* (eds Varoquaux, G., Vaught. T & Millman, J.) **2008**, 11–15 (SciPy).

19. Travis, E.; Oliphant, E. A guide to NumPy USA: *Trelgol Publishing*. **2006**

20. Gansner, E. R.; North, S. C. An open graph visualization system and its applications to software engineering *Pract. Exper*. **2000**, *30*, 1203– 1233.

21. Jurafsky, D.; Martin, J. H. Speech and language processing Copyright .All rights reserved. *Prentice Hall, Englewood Cliffs*, New Jersey, Draft of September 11 **2018**.

22. Schaftenaar, G.; Noordik, J. H. Molden: a pre- and post-processing program for molecular and electronic structures *J. Comput.-Aided Mol.* Design **2000**, *14*, 123-134.

23. Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics *J. Molec. Graphics* **1996**, *14.1*, 33-38.

24. Viterbi, A. J. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm *Transactions on Information Theory*. **1967**, *13*, 260–269.

25. Wang, L. P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martinez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem* **2014***, 6*, 1044-1048.

26. Ritson, D.; Sutherland, J. D. Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat Chem* **2012**, *4*, 895–899.

27. Eckhardt, A. K.; Linden, M. M.; Wende, R. C.; Bernhardt, B.; Schreiner, P. R. Gas-phase sugar formation using hydroxymethylene as the reactive formaldehyde isomer. *Nat* Chem **2018**, *10*, 1141-1147.